

Visualization of the operational space of edge-localized modes through low-dimensional embedding of probability distributions

A. Shabbir^{1,2, b)} G. Verdoolaege^{1,3}, O.J.W.F Kardaun², J.M. Noterdaeme^{1,2} and JET-EFDA Contributors*

JET-EFDA, Culham Science Centre, Abingdon OX14 3DB, UK

¹Department of Applied Physics, Ghent University, B-9000 Ghent, Belgium

²Max-Planck-Institut für Plasmaphysik, Garching D-85748, Germany

³Laboratorium voor Plasmafysica van de KMS, Koninklijke Militaire School, B-1000 Brussels, Belgium

(Presented XXXXX; received XXXXX; accepted XXXXX; published online XXXXX)

Information visualization aimed at facilitating human perception is an important tool for the interpretation of experiments on the basis of complex multidimensional data characterizing the operational space of fusion devices. This work describes a method for visualizing the operational space on a two-dimensional map and applies it to the discrimination of type I and type III edge-localized modes (ELMs) from a series of ELMy discharges at JET. The approach accounts for stochastic uncertainties that play an important role in fusion data sets, by modeling measurements with probability distributions in a metric space. The method is aimed at contributing to physical understanding of ELMs as well as their control. Furthermore, it is a general method that can be applied to the modeling of various other plasma phenomena as well.

I. INTRODUCTION

¹Visualization of the operational space of fusion devices is an important aid for exploratory and confirmatory analysis of the behavior of specific operational regimes and plasma instabilities. It entails the representation of multidimensional and often complex nonlinear experimental data models, describing the plasma state, in a two-dimensional diagram. This can provide physicists and machine operators with a convenient means and a useful tool for studying data patterns (relationships, clusters) reflecting key regimes and their underlying physics.

High-confinement tokamak plasmas are characterized by repetitive bursts called *edge-localized modes* (ELMs), which lead to partial collapse of the pedestal pressure and consequent exhaust of energy and particles towards the wall and divertor targets. Characteristics of ELMs, their control and physical understanding are important for ITER and next-step fusion devices. As a result, investigation into the precise effect of plasma parameters on the ELM behavior and type is crucial.

In this work, we present an information visualization tool for projecting high-dimensional plasma data into a two-dimensional (2D) map and apply it for detecting cluster structure corresponding to type I and type III ELMs in an assembled dataset of JET plasma discharges. Further, this work attempts to incorporate the substantial stochastic uncertainty affecting the ELM properties, by representing the measurements of each discharge by a probability distribution. Through a probabilistic description of plasma signals, our method tries to utilize the information content residing in the error bars associated with each

measurement alongside being potentially capable of taking into account higher-order statistical moments.

II. Visualization of data probability distributions

An important goal in information visualization is often to assess the proximity, or alternatively the dissimilarity, of data points. Multi-dimensional scaling (MDS)¹⁻², illustrated in FIG.1, is an established technique which yields a projection in the 2D Euclidean plane of high-dimensional data, while ensuring minimal information loss during dimensionality reduction.

Let $X = \{1, 2, \dots, N\}$ be a set of N high dimensional data points where the distance (or *dissimilarity*) between the i_{th} and j_{th} data point is d_{ij} . Given a pair-wise distance matrix, for high dimensional data points in X , MDS finds two (or possibly 3) dimensions for mapping the high dimensional data points such that $d'_{ij} \approx d_{ij}$ for all data points, where, d'_{ij} is the distance between i_{th} and j_{th} data point in two dimensions. Hence, the 2D maps resulting from MDS ensure minimal distortion of all pairwise distances between data points.

Data points in our context are probability distributions, with each

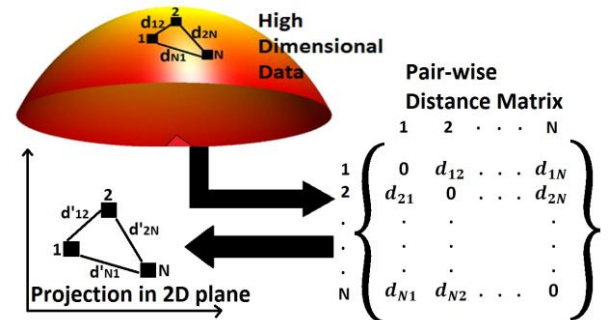


FIG. 1. Illustration of MDS. High dimensional data points are projected in 2 dimensions, such that $d'_{ij} \approx d_{ij}$ where d_{ij} and d'_{ij} are distances between data points i and j in high dimensions and in 2D respectively.

* See the Appendix of F.Romanelli et al., *Proceedings of the 24th IAEA Fusion Energy Conference 2012, San Diego, USA.*

^{a)} Contributed paper published as part of the *Proceedings of the 20th Topical Conference on High-Temperature Plasma Diagnostics, Atlanta, Georgia, June, 2014.*

^{b)} aqa.shabbir@ugent.be

plasma measurement being represented by a unique probability density function (PDF). Since *distance* is essentially a geometric concept, a description of the geometry of probability distributions is required. To this end, we employ the mathematical framework of *information geometry*, which treats a family of PDFs, e.g. the Gaussian family, as a space wherein each point represents a single PDF. The Fisher information provides a metric in this space, defining the concept of a distance between PDFs which is known as the *Rao geodesic distance* (GD). The GD is a natural and theoretically motivated measure between the PDFs^{3,4}. A closed-form expression for the GD, also, exists in the case of a univariate Gaussian model, $\mathcal{N}(\mu, \sigma)$, parameterized by its mean μ and standard deviation σ . This allows fast computation of the distance. Further, in the case of multiple independent Gaussian variables, the squared GD between two sets of products of distributions is given by the sum of the squared GDs between corresponding individual distributions^{2,3}.

In this paper we present MDS utilizing GDs between PDFs as an information visualization tool for yielding 2D maps for high-dimensional plasma data.

III. PLASMA PARAMETERS FOR EDGE-LOCALIZED MODES

Discrimination of various observed classes of ELMs is a significant challenge. The presented visualization methodology is employed here for discrimination of type I and type III ELMs from a series of carbon-wall JET plasmas between the years 2000 and 2009 with MarkGB and MarkGBRS as divertors. To this purpose, from the range of discharge numbers [50564, 76871], a database of 68 JET plasmas pertaining to type I ELMs, 26 JET plasmas of type III ELMs and 5 JET plasmas [66105-66109] of so-called type I high-frequency ELMs have been analyzed. This is an extension of the data set used earlier by Webster *et al.*⁵ for statistical characterization of ELM types. We call this dataset JET_CW_ELM database (DB1), henceforth referred as JET_CW_ELM (DB1). The analysis, in this work, has been restricted to time intervals in which the plasma conditions were quasi-stationary. Further, all experiments dealing with ELM control and mitigation techniques have been excluded so as to obtain the current version (v1) of the ‘standard subset’ of JET_CW_ELM (DB1).

A robust ELM algorithm was employed for the extraction of inter-ELM time intervals for each plasma discharge from the measured Balmer-alpha radiation signal from deuterium ($D\alpha$) at JET’s inner divertor. Gaussian probability density functions (PDFs) have then been used for capturing the statistics of inter-ELM waiting times, as shown in FIG 2.

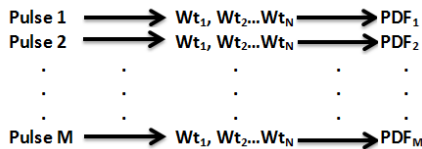


FIG. 3. Each pulse (plasma discharge) is represented as a series of waiting times, followed by modelling by a suitable probability distribution function (PDF), where there are M pulses and each pulse has N waiting times.

In addition, density-normalized input power ($\langle P_n \rangle$) [keV/s], normalized electron temperature ($\langle T_n \rangle$) [keV] and line-integrated edge electron density (n_e) [$10^{19} m^{-2}$], have also been included in the dataset. A probability distribution was used to model the data

corresponding to each plasma parameter. For simplicity it was assumed that the error bars associated with each plasma parameter pertain to a statistical uncertainty in the data, specifically that they represent a single standard deviation. Theoretically, the underlying probability distribution is Gaussian with mean the measurement itself and standard deviation the error bar. $\langle P_n \rangle$ and $\langle T_n \rangle$ are given as follows:

$$\langle P_n \rangle = \frac{1}{1.602 \times 10^3} * \frac{P_{input}}{n_{e,19,(vol.avg)} * Volume} \text{ keV/s} \quad (1)$$

where $P_{input} = P_{ohmic} + P_{NBI} + P_{ICRH}$, and

$$kT = \langle T_n \rangle = \frac{1}{1.602 \times 10^3} * \frac{1}{3} * \frac{W_{thermal}}{n_{e,19,(vol.avg)} * Volume} \text{ keV} \quad (2)$$

It should be stressed that each plasma parameter is a two-dimensional quantity with the mean and standard deviation of the respective probability distribution constituting the two dimensions.

IV. VISUAL MAPS

2D visual mapping for the operational space pertinent to type I and type III ELMs are presented in FIG.3 to FIG.6. In FIG. 3 to FIG.5, MDS is coupled with GD between the probability distributions for projecting the information. In FIG.3, inter-ELM waiting times, modelled with Gaussian PDFs, are projected and the ELM type for each pulse is indicated. The projection appears to be consistent, as the type I HF ELMs that have frequencies similar to type III ELMs are mapped in the proximity of type III ELMs. In addition, the type III ELM pulse 50567, which has a frequency ($f_{ELM} \approx 40$ Hz) more typical of type I ELMs, is projected close to the cluster of type I ELMs.

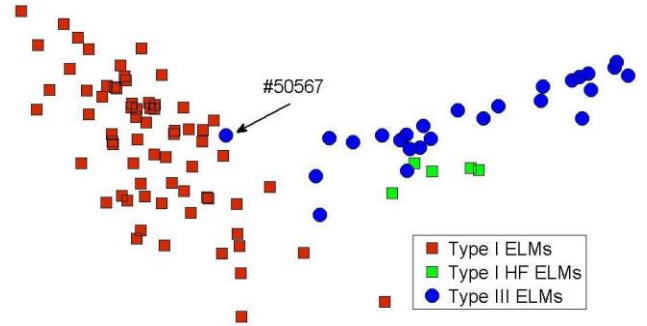


FIG. 2. Projection of inter-ELM time intervals using MDS with GD. ELM types are indicated.

FIG.4 explicitly tracks the change in ELM frequency and the consequent change in ELM type as one navigates through the map. FIG.4 is an indicator of the potential of these visual maps as they allow for tracking of changing values of a certain plasma parameter in the operational space. FIG.5a incorporates two additional parameters in the visual map, i.e. $\langle P_n \rangle$ and $\langle T_n \rangle$. The striking observation is that as more information is incorporated in the visualization, type I HF ELMs and type III pulse 50567 now lie with the clusters of type I and type III, respectively. Hence, incorporating additional global plasma parameters improves the accuracy of the map. FIG. 5b incorporates n_e as another additional parameter, and more correct placement of type I HF ELMs and pulse 50567 can also be observed here. In addition, the line of best separation between type I and type III ELMs is shown in FIG. 5a and FIG. 5b. This can aid in the classification of ELMs when the ELM type is unknown. Projection on this map of a new pulse for which the ELM type is unknown, can provide an indication of the ELM type.

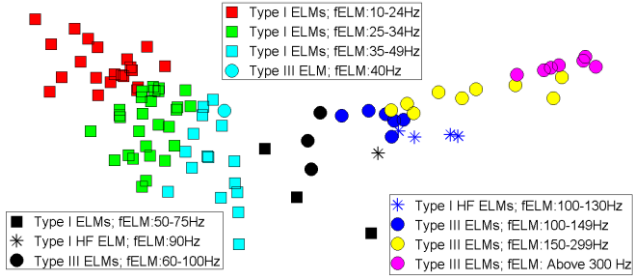
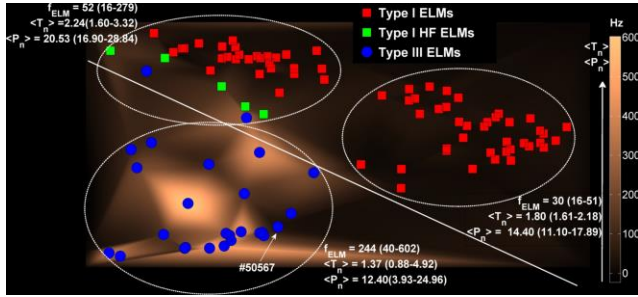
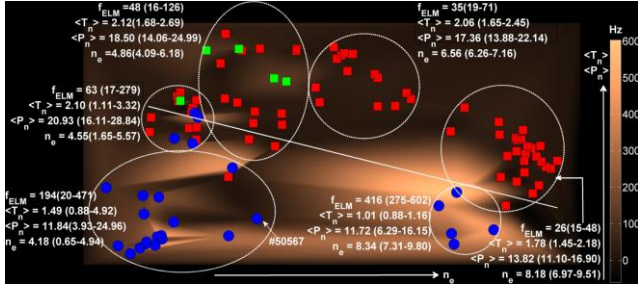


FIG. 4: Projection of inter-ELM time intervals using MDS with GD. The change in ELM frequency is tracked.



(a)



(b)

FIG. 5: Projections obtained using MDS with GD. (a) is a map of ELM waiting times, $\langle P_n \rangle$ and $\langle T_n \rangle$. (b) is a map of ELM waiting times, $\langle P_n \rangle$, $\langle T_n \rangle$ and n_e . Clusters are identified and the mean value for each parameter for the respective cluster is given. The range of the parameter values for pulses in each cluster is specified within brackets. The line of best separation between the type I and type III ELMs is also depicted. Trends in change in plasma parameters across the map are also highlighted.

In FIG.5a and FIG.5b cluster structure amongst the plasma discharges is identified. Further the trends in the changing values of plasma parameters along the clusters are also indicated. In both FIG. 5a and FIG. 5b $\langle P_n \rangle$ and $\langle T_n \rangle$ increase from the bottom of the map to the top, as a transition is made from the clusters of Type III ELMs to Type I ELMs. Background color gradation indicates the changing ELM frequency within the map.

Ordinary least squares linear regression using the two dimensions yielded by MDS (with GD) as predictors for plasma parameters which are successively regarded as the dependent variables yields appreciably good results. This provides an intuitive insight into the merit of the mappings as despite the lack of a direct relationship between the dimension coordinates and each of the plasma

parameters, the mappings prove to be reasonably reliable predictors of the plasma parameters as indicated by root mean square error (RMSE and R-squared values for the models. The results are given in table 1.

$$Model: \log y \sim 1 + x_1 + x_2 + x_1^2 + x_2^2$$

For example :

$$\log n_e = 1.77 - 117x_1 - 7.3x_2 + 5162x_1^2 + 6914x_2^2$$

Where, x_1 and x_2 are co-ordinates of the two dimensions yielded by MDS-GD

Dependent variable (logy)	R-squared	RMSE
$\langle T_n \rangle$	0.77	0.14
$\langle P_n \rangle$	0.76	0.14
n_e	0.61	0.20
f_{ELM}	0.51	0.66

Table 1: Regression for plasma parameters using the dimension coordinates yielded by MDS with GD as predictors

V. CONCLUSIONS AND OUTLOOK

An information visualization tool, that effectively incorporates measurement uncertainties, has been presented and its potential has been illustrated with an application to visualization of the operational space of different ELM types. The presented method is generic, potentially capable of visualizing multi-machine data in a single map and allows simultaneous viewing of numerous (more than three) plasma parameters on the same map. Further, it can be adapted with relative ease if the plasma parameters are best described by non-Gaussian PDF's such as Weibull or log-normal.

It can also be used for investigating the influence of plasma parameters on ELM characteristics. For example, the change of ELM types (type III to type I to type III) with input power within a single pulse can be mapped and monitored effectively. Furthermore, the developed tool can also potentially contribute to ELM control and mitigation, through visualization of e.g. ELM frequency and size distributions, and by quantifying the influence on these distributions of various control parameters.

ACKNOWLEDGEMENTS

This work was supported by EURATOM and carried out within the framework of the European Fusion Development Agreement. The views and opinions expressed herein do not necessarily reflect those of the European Commission.

¹T.F.Cox and M.A.A. Cox, "Multidimensional Scaling," Chapman and Hall, 2001.

²M.L. Davidson and S.G. Sireci, Chapter MDS in: H.E.A. Tinsley and S.D. Brown (Eds.): Handbook of applied multivariate statistics and mathematical modeling, Academic Press, 2000.

³G. Verdoolaege, G. Karagounis, M. Tendler and G. Van Oost, "Pattern recognition in probability spaces for visualization and identification of plasma confinement regimes and confinement time scaling," Plasma Phys. Control. Fusion **54**(12), 124006 (2012).

⁴G. Verdoolaege and P. Scheunders, "On the geometry of multivariate generalized Gaussian models," J. Math. Imaging. Vis., **43**(3), 180-193(2011).

⁵A.J.Webster and R.O.Dendy, "Statistical characterisation and classification of edge-localised plasma instabilities," Phys. Rev. Lett., **110**, 155004 (2013).